Cyclic LRC Codes and their Subfield Subcodes

Itzhak Tamo* Alexander Barg† Sreechakra Goparaju‡ Robert Calderbank§

***Abstract*—We consider linear cyclic codes with the locality property, or locally recoverable codes (LRC codes). A family of LRC codes that generalizes the classical construction of Reed-Solomon codes was constructed in a recent paper by I. Tamo and A. Barg (IEEE Trans. IT, no. 8, 2014). In this paper we focus on the optimal cyclic codes that arise from the general construction. We give a characterization of these codes in terms of their zeros, and observe that there are many equivalent ways of constructing optimal cyclic LRC codes over a given field. We also study subfield subcodes of cyclic LRC codes (BCH-like LRC codes) and establish several results about their locality and minimum distance.**

I. Introduction

Locally recoverable codes (LRC codes) have been extensively studied in recent literature following their introduction in [4]. A linear code $\mathcal{C} \subset \mathbb{F}_q^n$ is called locally recoverable with locality r if the value of every symbol of the codeword depends only on r other symbols of the same codeword. If $\dim \mathcal{C} = k$, then clearly $r \le k$. Applications of LRC codes in distributed storage motivate constructions in which r is a small constant, while n and k could be large. Early constructions of LRC codes such as [6], [8], [9], [10], [12] relied on alphabets of cardinality much greater than the code length. Paper [11] introduced a family of LRC codes of Reed-Solomon (RS) type over field alphabets of size comparable to the code length n. We call these codes RS-like codes below. Some of the codes constructed in [11] are cyclic of length $n|(q-1)$, where q is the size of the field. In this paper we focus on cyclic RS-like codes. As our first result, we characterize the distance and the locality parameter of such codes in terms of the code's zeros. We also study subfield subcodes of RS-like codes and describe the locality parameter in terms of irreducible cyclic codes supported on the coordinate subsets that form the recovering sets of the original code. This enables us to find estimates of the locality parameter based on the structure of the zeros of the code and to construct examples of binary LRC codes.

The general question of finding the locality r is equivalent to finding the dual distance of a cyclic code, which is a difficult problem. However unlike for the problem of error correction, we actually gain by proving that the dual distance is smaller than the estimated value, as this implies better local recovery properties of the LRC code. Subfield subcodes are particularly fascinating as they not only increase the distance, but also reduce the locality, though at the expense of code dimension.

Apart from [11], the paper particularly relevant to this study is [5]. In it, the authors construct several examples of binary cyclic LRC codes with locality 2 and in a number of cases prove optimality of their constructions.

The following Singleton-like bound on the distance d of an (n,k,r) LRC code was proved in [4]: $d \le n-k-\lceil k/r \rceil + 2$. We call the code optimal if its distance meets this bound with equality.

II. The Reed-Solomon-like construction

Let us briefly recall the construction detailed in [11]. Our aim is to construct an LRC code over $\mathbb{F}_q$ with the parameters (n,k,r), where $n \le q$. We additionally assume that $(r+1)|n$ and $r|k$, although both the constraints can be lifted by adjustments to the construction presented below [11]. Throughout this paper we let

$$\nu = n/(r+1),\ \mu = k/r.$$

Let $p(x) \in \mathbb{F}_q[x]$ be a polynomial of degree $r+1$ such that there exists a partition $\mathcal{A} = \{A_1, \dots, A_\nu\}$ of a set of points $A = \{P_1, \dots, P_n\} \subset \mathbb{F}_q$ into subsets of size $r+1$ such that $p(x)$ is constant on each set $A_i \in \mathcal{A}$.

Consider the k-dimensional linear subspace $V \subset \mathbb{F}_q[x]$ spanned by the set of k polynomials

$$\{p(x)^j x^i,\ i=0,\dots,r-1; j=0,\dots,\mu-1\}. \tag{1}$$

Given an information vector $a = (a_{ij}, i=0,\dots,r-1; j=0,\dots,\mu-1) \in \mathbb{F}_q^k$ let

$$f_a(x) = \sum_{i=0}^{r-1} \sum_{j=0}^{\mu-1} a_{ij} p(x)^j x^i. \tag{2}$$

Note that $f_a(x)$ belongs to the subspace V. Now define the code $\mathcal{C}$ as the image of the linear evaluation map

$$\begin{aligned} e : V &\to \mathbb{F}_q^n \\ f_a &\mapsto (f_a(P_i), i=1,\dots,n). \end{aligned} \tag{3}$$

The minimum distance of the code $\mathcal{C}$ equals $d = n - k(r+1)/r + 2$, and is optimal for the given parameters. The code also has the LRC property: namely, the value of the symbol in coordinate $P \in A_i \in \mathcal{A}$ can be found by interpolating a polynomial of degree $\le r-1$ that matches the codeword at the points $P_j \in A_i \backslash \{P\}$. Below we call the subset of coordinates $A_i \backslash \{P\}$ the *recovering set* of the coordinate P.

III. Cyclic q-ary LRC codes

In this paper we are concerned with the following special case of the construction (2)-(3). Let $n|(q-1)$ and choose the polynomial $p(x)$ in (1) to be the annihilator polynomial of a subgroup of the multiplicative group $\mathbb{F}_q^*$. As shown in [11], the polynomial f_a in (2) can be taken in the form

$$f_a(x) = \sum_{\substack{i=0 \\ i \ne r \bmod (r+1)}}^{\mu(r+1)-2} a_i x^i. \tag{4}$$

[*] I. Tamo is with the Dept. of EE-Systems, Tel Aviv University, Tel Aviv, Israel. The research was done while at the Institute for Systems Research, University of Maryland, College Park, MD 20742 (email: zactamo@gmail.com).

[†] A. Barg is with the Dept. of ECE and ISR, University of Maryland, College Park, MD 20742 and IITP, Russian Academy of Sciences, Moscow, Russia (email: abarg@umd.edu). Research of A. Barg and I. Tamo supported by NSF grants CCF1422955, CCF1217894, and CCF1217245.

[‡] S. Goparaju is with CALIT2, University of California, San Diego, CA 92093 (email: sgoparaju@ucsd.edu).

[§] R. Calderbank is with the Dept. of ECE, Duke University, NC 27708 (email: robert.calderbank@duke.edu).

Choose the set of evaluation points as $A = \{1, \alpha^1, \dots, \alpha^{n-1}\}$, where α is a primitive n-th root of unity, and construct a linear code $\mathcal{C}$ using the evaluation map (3).

Using this representation as the starting point, we observe that $\mathcal{C}$ is a cyclic code of length n. Generally, a cyclic code is an ideal in the ring $\mathbb{F}_q[x]/(x^n - 1)$ which is generated by a polynomial $g(x)$ such that $g(x)|(x^n - 1)$. Let $\mathbb{F}_{q^m}$ be an extension field that contains the n-th roots of unity. Let $t = \deg(g)$ and let $Z = \{\alpha^{i_j}, j = 1, \dots, t\} \subset \mathbb{F}_{q^m}$ be the zeros of $g(x)$. The set of unique representatives of cyclotomic cosets in Z with respect to the field $\mathbb{F}_q$ is called a *defining set* of zeros of the code $\mathcal{C} = \langle g(x) \rangle$. Throughout this section we assume that $m = 1$, i.e., that $n|(q-1)$, each cyclotomic coset is of size one, and the defining set is Z.

As our first result in this section, we identify the zeros of the code $\mathcal{C}$ constructed using representation (4). Next we make some observations regarding the structure of zeros of cyclic LRC codes. Based on these, we introduce a general construction of optimal q-ary cyclic codes, described in the following theorem.

Theorem 3.1: Let α be a primitive n-th root of unity, where $n|(q-1)$; $l, 0 \le l \le r$ be an integer; and $b \ge 1$ be an integer such that $(b, n) = 1$. Let $\mu = k/r$. Consider the following sets of elements of $\mathbb{F}_q$:

$L = \{\alpha^i, i \bmod (r+1) = l\}$, and

$D = \{\alpha^{j+sb}, s = 0, \dots, n - \mu(r+1)\}$,

where $\alpha^j \in L$. The cyclic code with the defining set of zeros $L \cup D$ is an optimal (n, k, r) q-ary cyclic LRC code.

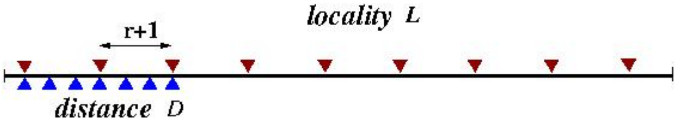

Fig. 1: Subsets of zeros for distance (D) and locality (L).

It will be seen that the set D accounts for the code's distance, while L ensures the locality property.

The proof of this theorem follows from Lemmas 3.2 and 3.3 and is given at the end of this section. Recall the following property where α is an n-th root of unity and p is the characteristic of the field:

$$\sum_{i=0}^{n-1} \alpha^i = \begin{cases} n \bmod p, & \text{if } \alpha = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Lemma 3.2: Consider the cyclic code $\mathcal{C}$ of length n constructed using the polynomials $f_a(x)$ given by (4). The rows of the generator matrix $\mathcal{G}$ of $\mathcal{C}$ have the form $(1, \alpha^j, \alpha^{2j}, \dots, \alpha^{(n-1)j})$, for all j such that

$j \in \{0, 1, \dots, \mu(r+1) - 2\} \backslash \{s(r+1) - 1, s = 1, \dots, \mu - 1\}$.

The defining set of zeros of $\mathcal{C}$ has the form $R = D \cup \bar{L}$, where

$$D = \{\alpha^i : i = 1, ..., n - \mu(r+1) + 1\}$$
$$\bar{L} = \{\alpha^{n-(\mu-l)(r+1)+1},\ l = 1, 2, \dots, \mu - 1\}$$

The code $\mathcal{C}$ is an optimal (n, k, r) LRC code with distance $d = n - \mu(r+1) + 2$.

Proof: The statement about the generator matrix follows directly from (4). To prove the statement about the zeros, it suffices to show that the dot product of any row of $\mathcal{G}$ and the row vector $(1, \alpha^t, \alpha^{2t}, ..., \alpha^{(n-1)t})$ for any $t \in R$ is zero. Indeed, from (5), if α^j is the generating element of a row of $\mathcal{G}$ and $t \in R$, we need to show that $\alpha^{j+t} \ne 1$, or that $j + t$ is not a multiple of n. This is true because if $t \in D$, then $j + t \le n - 1$, and if $t \in \bar{L}$, then

$$j + t = n - ((\mu - l)(r+1)) + 1 + j, \tag{6}$$

where $l = 1, 2, \dots, \mu - 1$. The first two terms on the RHS of (6) are multiples of $r + 1$, therefore the entire RHS is a multiple of $r + 1$ if and only if so is $j + 1$. Since $\mathcal{G}$ does not include the rows that would make the latter possible, we have $(r+1) \not| (j + t)$. Finally, the claim about the distance follows from the BCH bound on the set of zeros D. ■

In Lemma 3.2, we described the set of zeros of $\mathcal{C}$ as a union of two disjoint subsets of roots of unity. Alternatively, the set of exponents R obviously can be described as a union of two non-disjoint sets, $R = D \cup L$, where D is as given in Lemma 3.2 and

$$L = \{\alpha^{j(r+1)+1}, j = 0, 1, \dots, \nu - 1\}.$$

As already observed, the subset D guarantees a large value of the code distance, supporting the optimality claim. It is natural to assume that the zeros in L account for the locality property. The following lemma shows that this is indeed the case.

Lemma 3.3: Let $0 \le l \le r$ and consider a $\nu \times n$ matrix $\mathcal{H}$ with the rows

$$h_m = (1, \alpha^{m(r+1)+l}, \alpha^{2(m(r+1)+l)}, \dots, \alpha^{(n-1)(m(r+1)+l)}),$$

where $m = 0, 1, \dots, \nu - 1$, and $\nu = n/(r+1)$. Then all the cyclic shifts of the n-dimensional vector of weight $r + 1$

$$v = (1 \underbrace{0 \dots 0}_{\nu-1} \alpha^{l\nu} \underbrace{0 \dots 0}_{\nu-1} \alpha^{2l\nu} \underbrace{0 \dots 0}_{\nu-1} \dots \alpha^{rl\nu} \underbrace{0 \dots 0}_{\nu-1})$$

are contained in the row space of $\mathcal{H}$.

Proof: First note that $av = \sum_{m=0}^{\nu-1} h_m$, where $a = \nu \bmod p$. Indeed,

$$\sum_{m=0}^{\nu-1} \alpha^{j(m(r+1)+l)} = \alpha^{lj} \sum_{m=0}^{\nu-1} (\alpha^{j(r+1)})^m.$$

The element $\alpha^{j(r+1)}$ is a ν-th root of unity, so by (5) the last sum is zero if j is not a multiple of ν and $a\alpha^{lj}$ otherwise. We conclude that the vector av is contained in the row space of $\mathcal{H}$, and since $a \in \mathbb{F}_q, a \ne 0$ so is the vector v itself. The row space of $\mathcal{H}$ over $\mathbb{F}_q$ is closed under cyclic shifts, and this proves the lemma. ■

Note that $\mathcal{H}$ forms a parity-check matrix of the code with defining set $Z_l = \alpha^l \cdot \{\alpha^{m(r+1)}, m = 0, 1, \dots, \nu - 1\}, 0 \le l \le r$. The cyclic shifts of the vector v partition the support of the code into disjoint subsets of size $r+1$ which define the local recovering sets of the symbols. Therefore we obtain the following statement.

Proposition 3.4: Let $\mathcal{C}$ be a cyclic code of length n over $\mathbb{F}_q$ with the complete defining set Z, and let r be a positive integer such that $(r+1)|n$. If Z contains some coset of the group of ν-th roots of unity, then $\mathcal{C}$ has locality at most r.

Remark 1: Lemma 3.3 provides a general method of constructing optimal cyclic q-ary linear codes. The construction is rather flexible and relies on the choice of two sets of zeros of the code, D and L, which are responsible for error correction capability and locality of $\mathcal{C}$. In other words, the set D accounts for the distance properties of the code while L takes care of the locality property. The possibility to shift L and D around will prove useful in the next section where it will enable us to improve the locality of subfield subcodes of our codes.

Remark 2: In [11] it was also observed that the construction (2)-(3) can be used to construct codes with two (or more) disjoint recovering sets for every symbol of the encoding. Turning to cyclic codes, we note that Proposition 3.4 provides a simple sufficient condition for such a code to have several recovering sets: all we need is that the complete defining set contain cosets of subgroups of groups of unity of degree $\nu_1, \nu_2, \dots,$ where the ν_i's are pairwise coprime. For instance a cyclic code of length $n = 63$ whose complete defining set contains the sets of 7-th and 9-th roots of unity, has two *disjoint* recovering sets of sizes 6 and 8 for every symbol.

We conclude by proving the main result of this section.

Proof of Theorem 3.1: The minimum distance of the code $\mathcal{C}$ is estimated from below using the BCH bound for the set of zeros D. That the locality parameter equals r follows from Proposition 3.4 used for the set L. The dimension of the code equals $n - |D \cup L| = k$. This completes the proof. ■

IV. Subfield Subcodes

A large part of the classical theory of cyclic codes is concerned with subfield subcodes of Reed-Solomon codes, i.e., the BCH codes, and related code families. In this section we pursue a similar line of inquiry with respect to cyclic LRC codes introduced in the previous section. In particular, through an analysis of parameters of the BCH-like codes and some examples, we derive stronger bounds on locality with the same set of zeros L that we considered in the previous section.

A. Notation

Let Z be the complete defining set of the code $\mathcal{C}'$ over $\mathbb{F}_q$, (i.e., a BCH-type code) and let $\mathcal{C}$ the corresponding Reed-Solomon type code, i.e., the cyclic code over $\mathbb{F}_{q^m}$ with the same set of zeros. In the previous section we considered cyclic codes where the symbol field and the locator field coincided, as is common for Reed-Solomon codes. In the context of subfield subcodes, the symbol field will be denoted $\mathbb{F}_q$ and the locator field $\mathbb{F}_{q^m}$ (for most of our examples, $q = 2$). The field $\mathbb{F}_{q^m}$ is the splitting field of the generator polynomial $g(x)$, while over $\mathbb{F}_q$ we have $g(x) = \prod_{j\in J} m_{i_j}(x)$, where $(i_j, j \in J)$ is the set of representatives of the cyclotomic cosets that form the defining set of zeros of $\mathcal{C}$, and m_{i_j}'s are the corresponding minimal polynomials.

Given a code $\mathcal{C} \subset \mathbb{F}_{q^m}^n$, its *subfield subcode* $\mathcal{C}' = \mathcal{C}_{|\mathbb{F}_q}$ consists of the codewords of $\mathcal{C}$ all of whose coordinates are in $\mathbb{F}_q$. For the analysis of subfield subcodes we will use the trace mapping T_m from $\mathbb{F}_{q^m}$ to $\mathbb{F}_q$, defined as

$$T_m(x) = x + x^q + ... + x^{q^{m-1}}, x \in \mathbb{F}_{q^m}.$$

Given a vector $v = (v_1, \dots, v_n) \in \mathbb{F}_{q^m}^n$, we use the notation $T_m(v) := (T_m(v_1), \dots, T_m(v_n))$. The trace of the code $\mathcal{C} \subset \mathbb{F}_{q^m}^n$ is the code over $\mathbb{F}_q$ obtained by computing the trace of all vectors $c \in \mathcal{C}$, i.e.,

$$T_m(\mathcal{C}) = \{T_m(c), c \in \mathcal{C}\}.$$

Let $\mathcal{C}^\perp$ be the dual code of a cyclic code $\mathcal{C}$. Obviously, the locality parameter $r(\mathcal{C})$ equals the dual distance $d^\perp(\mathcal{C}) := d(\mathcal{C}^\perp)$. The dual code of the subfield subcode is characterized by *Delsarte's Theorem.*

Theorem 4.1: [2, Theorem 2] The dual of a subfield subcode is the trace of the dual of the original code, i.e., $(\mathcal{C}_{|\mathbb{F}_q})^\perp = T_m(\mathcal{C}^\perp)$.

Remark: If $\mathcal{C}$ is an (n, k, r) LRC code, then any coordinate in the dual code is contained in the support of a codevector of weight at most $r + 1$. Hence by Theorem 4.1, the subfield subcode $\mathcal{C}_{|\mathbb{F}_q}$ has locality $\le r$. This observation is not surprising since the trace mapping T_m does not increase the weight of a codeword. However, as we shall show in the sequel, the locality can be, and in most cases is, much smaller than r.

B. Preliminaries: From locality to irreducible cyclic codes

Let $\mathcal{C}'$ and $\mathcal{C}$ be the codes defined in Section IV-A. Proposition 3.4 states that if Z contains some coset $\{\alpha^i : i \bmod (r + 1) = l\}$ of the subgroup generated by α^{r+1} then $\mathcal{C}$ has locality r. By Lemma 3.3, the dual code $\mathcal{C}^\perp$ contains the vector

$$v = (1\,\underbrace{0...0}_{\nu-1}\beta^l\,\underbrace{0...0}_{\nu-1}\beta^{2l}\,\underbrace{0...0}_{\nu-1}\beta^{3l}\,\underbrace{0...0}_{\nu-1}...\beta^{rl}\,\underbrace{0...0}_{\nu-1}) \qquad (7)$$

where $\beta = \alpha^\nu$ is a primitive root of unity of degree $r+1$. The weight of the vector v is $\mathrm{wt}_H(v) = r + 1$ and the supports of its cyclic shifts partition the set of n coordinates of the code into subsets of size $r+1$. As noted above, these subsets define the local recovering sets A_i for the code $\mathcal{C}$. By Theorem 4.1, for any $\gamma \in \mathbb{F}_{q^m}$ and $v \in \mathcal{C}^\perp$, the vector $y := T_m(\gamma v) \in \mathcal{C}^\perp_{|\mathbb{F}_q} = (\mathcal{C}')^\perp$. Furthermore, $\mathrm{wt}_H(y) \le r + 1$, and if $y \ne 0$, then its nonzero coordinates form a recovering set of relatively small size in the code $\mathcal{C}'$.

In our analysis of the locality of the code $\mathcal{C}'$ we will restrict our attention to the following subspace of the code $(\mathcal{C}')^\perp$:

$$V = \langle T_m(\gamma v), \gamma \in \mathbb{F}_{q^m}\rangle. \qquad (8)$$

Below we make the following simplification. It will suffice to analyze only the nonzero coordinates of the subspace V, therefore, we will drop the zeros and treat v and all the derived vectors as vectors of length $r+1$ in $\mathbb{F}_{q^m}$ or $\mathbb{F}_q$, as appropriate. By abuse of notation, we still use the same letter v, and from now on write

$$v = (1, \beta^l, \beta^{2l}, ..., \beta^{rl}). \qquad (9)$$

Note that since below we rely only on a subset of the vectors in $(\mathcal{C}')^\perp$, the code $\mathcal{C}'$ might have a better (i.e., smaller) locality parameter than the one guaranteed by our results.

The form of the vectors in the subspace V (8) is reminiscent of the representation of vectors in irreducible cyclic codes [7], [13]. In this section we take this as a starting point, connecting locality and results about such codes.

Recall that a q-ary linear cyclic code is called *irreducible* if it forms a minimal ideal in the ring $\mathbb{F}_q[x]/(x^n - 1)$. The main result about irreducible codes is given in the following theorem.

Theorem 4.2: [13, Theorem 6.5.1] Let $s > 0$ be an integer, $m = \mathrm{ord}_s(q)$ be the multiplicative order of q modulo s, let β be a primitive s-th root of unity in $\mathbb{F}_{q^m}$. The set of vectors

$$V = \{(T_m(\gamma), T_m(\gamma\beta), \dots, T_m(\gamma\beta^{s-1}) : \gamma \in \mathbb{F}_{q^m}\}, \qquad (10)$$

is a $[s, m]$ linear irreducible code over $\mathbb{F}_q$. □

TABLE I
SOME EXAMPLES OF BINARY CODES FOR WHICH PROPOSITION 4.4 GIVES A TIGHT BOUND ON LOCALITY.[1]

n	k	d	$Z(\mathcal{C}')$	coset	z	r	w	$Z((\mathcal{C}')^{\perp})$	$d^{\perp}$	SH (11)	LP (12)	locator field $\mathbb{F}_{q^m}$
35	20	3	$\{1,15\}$	αG_7	3	$r \leq 3$	4	$\{0,1,7,15\}$	4	$k \leq 25$	$k \leq 29$	$\mathbb{F}_{2^{12}}$
45	33	3	$\{1\}$	αG_{15}	4	$r \leq 7$	8	$\{0,1,3,5,9,15,21\}$	8	$k \leq 37$	$k \leq 39$	$\mathbb{F}_{2^{12}}$
27	7	6	$\{1,9\}$	αG_3	2	$r = 1$	2	$\{0,3\}$	2			$\mathbb{F}_{2^{18}}$
63	36	3	$\{1,9,11,15,23\}$	αG_7	3	$r \leq 3$	4	$\{0,1,7,9,11,15,21,23\}$	4			$\mathbb{F}_{2^6}$

In the table, $Z(\mathcal{C})$ refers to the defining set of $\mathcal{C}$ (for brevity we write i instead of α^i); α is the n-th root of unity $\mathbb{F}_{q^m}$; w is the number of recovering sets A_i; other parameters are as given in Prop. 4.4. The columns labelled SH and LP refer to the bounds on LRC codes given in Appendix A.

Note that if in (10) we omit the requirement that β is a *primitive* root of unity, taking instead an s-th root of unity such that $\beta^t = 1$ for some $t|s$, then construction (10) results in a *degenerate* cyclic code. As is easily seen, in this case the code V consists of s/t repetitions of the irreducible code

$$\{(T_m(\gamma), T_m(\gamma\beta), ..., T_m(\gamma\beta^{t-1}) : \gamma \in \mathbb{F}_{q^m}\}.$$

C. The case $l = 0$

In this case we study a particular case of the above construction, taking $l = 0$ in (9). Then the complete defining set Z of the code contains the subgroup $G_{r+1} := \langle \alpha^{r+1} \rangle$ generated by the element α^{r+1} and we obtain $v = 1^{r+1}$ (the all-ones vector). By Theorem 4.2 the subspace V is of dimension 1 and is spanned by the all ones vector. Therefore the dual code $(\mathcal{C}')^{\perp}$ contains a vector of weight equal to $r+1$, which means that $\mathcal{C}'$ has the same recovering sets as the code $\mathcal{C}$.

Note that the subgroup $G_{r+1} = \{1, \alpha^{r+1}, \dots, \alpha^{r\nu}\}$ is closed under the Frobenius map, i.e.,

$$\forall_{\beta \in G_{r+1}} \; (\beta \in \; G_{r+1}) \; \Rightarrow \; (\beta^q \in \; G_{r+1}).$$

In other words, the set G_{r+1} is a union of cyclotomic cosets. Hence a cyclic code over $\mathbb{F}_q$ whose set of zeros contains G_{r+1} has the LRC property and is of large dimension.

Example 1: Let $\mathcal{C}'$ be a $[n = 45, k = 30, d = 4]$ binary cyclic code with zeros $\{0, 3, 5, 9\}$ in the field $\mathbb{F}_{2^{12}}$. Since the set of roots contains the subgroup G_9, we have $d^{\perp} \leq 9$, and hence the locality parameter of $\mathcal{C}$ satisfies $r \leq 8$; see (7). On the other hand, $(\mathcal{C}')^{\perp}$ has a defining set $\{1, 3, 7, 15\}$ and the parameters $[n = 45, k = 15, d = 9]$, so the value r is indeed 8.

To compare the parameters of this code with the upper bounds, we note that the shortening bound (SH) (11) gives $k \leq 3 \cdot 8 + k_2(45 - 3 \cdot 9, 4) = 36$. The linear programming bound (LP) (12) gives an estimate $M_2^{(c)}(45, 4, 8) \leq 2^{38.48}$ which translates into $k \leq 38$ (cf. Appendix A).
In this example the locality value predicted by our analysis is exact. This is not always the case as shown in the next example in which the locality is smaller than given by the estimate based on the vector v.

Example 2: Let $\mathcal{C}'$ be an $[21, 12, 4]$ binary cyclic code defined by the set of roots $\{0, 1, 7\}$ in $\mathbb{F}_{2^6}$. Since the set of roots contains the subgroup $\langle \alpha^7 \rangle$, the dual code has minimum distance at most 7, and hence the code has locality $r \leq 6$. On the other hand, $(\mathcal{C}')^{\perp}$ is a $[21, 9, 6]$ cyclic code with defining set $\{1, 3, 9\}$. Therefore the locality of $\mathcal{C}'$ is actually 5. From (11) and (12) we obtain, respectively, $k \leq 14$ and $k \leq 15$.

D. The case $l > 0$

The analysis of locality becomes more interesting if we take $l > 0$ in (9). Here we rely on the full power of the theory of irreducible cyclic codes, invoking several results that follow from the classical connection between these codes and Gauss sums. There are two options, namely $\gcd(l, r+1) = 1$ and $\gcd(l, r+1) > 1$. In the latter case, the analysis is as in the former except that we get a degenerate cyclic code. Below, if not stated, we exemplify the case $l > 0$ by taking $l = 1$.

Theorem 4.3: [3, Theorem 15] Consider a q-ary irreducible cyclic code V of length s as given in (10), where β and m are defined accordingly. Let $N = (q^m - 1)/t$ and assume that $\gcd(\frac{q^m-1}{q-1}, N) = 1$. Then V is a constant weight code over $\mathbb{F}_q$ of weight $(q-1)q^{m-1}/N$. □

If $q = 2$, the code V is the familiar simplex, or Hadamard, code of length $t = 2^m - 1$, dimension m and minimum distance $d = 2^{m-1}$. This follows since $Nt = 2^m - 1$ and $\gcd(2^m - 1, N) = 1$, and so $N = 1$. This leads to the following result.

Proposition 4.4: Let $z \geq 1$ be an integer such that $(2^z - 1)|n$ and let α be an n-th root of unity. Let $\mathcal{C}$ be an $[n, k]$ binary linear cyclic code whose complete defining set Z contains the coset αG_{2^z-1} of the group $G_{2^z-1} = \langle \alpha^{2^z-1} \rangle$. Then $\mathcal{C}$ has locality $r \leq 2^{z-1} - 1$. Moreover, each symbol of the code has at least 2^{z-1} recovering sets A_i of size $2^{z-1} - 1$.

Proof: Call V as V_m when defined using $\gamma \in \mathbb{F}_{q^m}$ and T_m. Note that $s = 2^z - 1|n$ and $n|2^m - 1$. The complete proof, relegated to Appendix B, uses the facts that β is an s-th root of unity in $\mathbb{F}_{q^z}$ (and so, also in $\mathbb{F}_{q^m}$), and that $V_m = V_z$. ■

Table I shows a few examples where an $[n, k, d]$ binary cyclic code $\mathcal{C}'$ with a defining set given by Z, contains the coset αG_{2^z-1}, and the upper bound on r obtained in Proposition 4.4 is tight. The last two codes in the table have dimensions far away from the bounds given in Appendix A.

Notice that for binary cyclic codes, when $l > 0$, we were able to reduce the upper bound on r roughly by a factor of 2 when the coset of a group G_s is contained in the defining set Z, where $s = 2^z - 1$. We show that this can be generalized to a q-ary cyclic code (the bound reduces roughly by a factor of $(q-1)/q)$) by a simple averaging argument to upper bound the distance of irreducible codes.

Proposition 4.5: Let V be a q-ary $[s, m, d]$ irreducible cyclic code, then its minimum distance satisfies $d \leq s(1 - \frac{q^{m-1}-1}{q^m-1})$.

Proof: For any element $\gamma \in \mathbb{F}_{q^m}$ define the linear mapping $T_{m,\gamma} : \mathbb{F}_q^m \to \mathbb{F}_q$ as $\alpha \mapsto T_m(\gamma\alpha)$, where the field $\mathbb{F}_{q^m}$ is viewed as a m dimensional vector space over $\mathbb{F}_q$. It is well known that these q^m linear mappings exhaust the set of all linear mappings. In other words, for any $\gamma \in \mathbb{F}_{q^m}$ there exists a vector $v_\gamma \in \mathbb{F}_q^m$ such that the mapping $T_{m,\gamma}$ is simply the scalar product with v_γ, i.e.,

$$T_{m,\gamma}(\alpha) = \langle v_\gamma, \alpha \rangle \text{ for any } \alpha \in \mathbb{F}_q^m.$$

Take a random *nonzero* mapping $T_{m,\gamma}$ and consider the set of indicator random variables $X_i = \mathbb{1}(T_{m,\gamma}(\beta^i) = 0), i =$

$0,...,s-1$. We have

$$P(X_i=1)\geq \frac{q^{m-1}-1}{q^m-1},$$

so $E|\{i:X_i=1\}|\geq s\frac{q^{m-1}-1}{q^m-1}$. We conclude that there exists a $\gamma\in\mathbb{F}_{q^m}$ such that weight of the codeword

$$\mathrm{wt}_H(T_m(\gamma),T_m(\gamma\cdot\beta),...,T_m(\gamma\cdot\beta^{s-1}))\leq s\Big(1-\frac{q^{m-1}-1}{q^m-1}\Big),$$

and the result follows. ■

Observe that this bound is tight for the simplex code.

Proposition 4.6: Let $\mathcal{C}$ be an $[n,k]$ a cyclic code over $\mathbb{F}_q$ such that its complete defining set contains the coset αG_s, where α is a primitive n-th root of unity and $s|n$, then the locality of $\mathcal{C}$ satisfies

$$r<s\Big(1-\frac{q^{m-1}-1}{q^m-1}\Big),$$

where m is the multiplicative order of q modulo s.

The theory of irreducible codes has been extensively explored, and for some cases their weight distribution is completely characterized. The technique behind these results is related to Gaussian sums and Gaussian periods [7]. We now cite a known result on irreducible codes, and cast it in the context of LRC codes. Observe that the upper bound on locality is again lower than that given by Proposition 3.4.

Theorem 4.7: [3, Theorem 17] Let $N=(q^m-1)/t$ and $\gcd(\frac{q^m-1}{q-1},N)=2$, then V is a two-weight code of length t and dimension m whose nonzero weights are $(q-1)(q^m\pm q^{m/2})/Nq)]$, and there are $(q^m-1)/2$ codewords of each of these weights.

Proposition 4.8: Let $\mathcal{C}'$ be an $[n,k]$ ternary cyclic code whose complete defining set Z contains the coset αG_t for some integer t that divides n, where α is an n-th root of unity. Let $N=(3^m-1)/t$, where $m=\mathrm{ord}_3\,(t)$. Assume that $\gcd(\frac{3^m-1}{2},N)=2$, then each symbol of the code $\mathcal{C}'$ has at least $3^{m-1}-3^{\frac{m}{2}-1}$ recovering sets of size less than $\frac{2(3^m-3^{\frac{m}{2}})}{3N}$.

Proof: The complete defining set Z of the code $\mathcal{C}'$ contains the set of roots αG_t, hence by Theorem 4.1 and (10), the $[n=(3^m-1)/N,k=m]$ irreducible cyclic code V is a shortened code of $(\mathcal{C}')^{\perp}$. By Theorem 4.7, the code V contains $\frac{3^m-1}{2}$ codewords of weight $(2(3^m-3^{\frac{m}{2}}))/3N$. Since the code is cyclic, each of its coordinates appears equally often as a nonzero coordinate of these codewords. Hence each coordinate of the code is nonzero in exactly $3^{m-1}-3^{\frac{m}{2}-1}$ codewords of weight $\frac{2(3^m-3^{\frac{m}{2}})}{3N}$ and the result follows. ■

Example 3: Let $\mathcal{C}'$ be a ternary cyclic code of length $n=80$ defined by the set of zeros $\{1,2,41\}$. Since each of the corresponding cyclotomic cosets is of size 4, the dimension of the code is $k=68$. The set of zeros contains α,α^{41}, so taking $t=40$ in Proposition 4.8 we obtain that $m=4$, and $d^{\perp}\leq 24$. Furthermore, each symbol of the code has at least 24 recovering sets of size 23.

For completeness, we present an example where $l\neq 1$.

Example 4: Let $\mathcal{C}$ be an $[63,54,2]$ binary cyclic code with the defining set $\{3,27\}$. In this case the complete defining set contains the coset α^3G_{21}, where α is a primitive root of unity of degree 63. Further, note that $\gcd(3,21)>1$, so the subcode V of $\mathcal{C}^{\perp}$ is a triple repetition of the $[7,3,4]$ simplex code. Therefore, the minimum distance of $\mathcal{C}^{\perp}$ is at most $3\cdot 4=12$ and the locality $r\leq 11$. It can in fact be shown that $\mathcal{C}^{\perp}$ is an $[63,9,12]$ cyclic code, so $r=11$.

E. Multiple Recovering Sets

Proposition 4.4 shows that each symbol has several recovering sets. Apart from the number of these sets, their structure is also of importance. For instance, we would like to know whether a symbol has a pair of *disjoint* recovering sets, which allows a parallel independent recovery of the lost symbol. While not a complete answer, we provide some analysis below. Recall that in Proposition 4.4, the subcode V of $\mathcal{C}^{\perp}$ is the simplex code. Consider $S_i\subseteq[t]$ a support of some codeword of V. By considering the generator matrix of V it is clear that S_i corresponds to an affine space defined by a vector in $u_i\in\mathbb{F}_2^z$. This observation yields a formula for size of the intersection of the supports of codewords of V.

Proposition 4.9: Let $S_i, i\in I$ be the supports of a subset of codewords in V. Then the size of the intersection

$$|\cap_{i\in I}S_i|\leq 2^{z-\mathrm{rk}(u_i,i\in I)}.$$

Proof: It can be easily checked that the set of vectors that contribute to the LHS is the set of all vectors $x\in\mathbb{F}_2^z$ that are a solution for the set of linear non-homogeneous equations $x\cdot u_i=1$, and the result follows. ■

For instance, for the $[63,36,3]$ code given in Table I, Proposition 4.9 gives tight bounds; we have $z=3$, and any two recovering sets of a symbol intersect in exactly one coordinate, while the intersection of any three is empty.

REFERENCES

[1] V. Cadambe and A. Mazumdar, *An upper bound on the size of locally recoverable codes*, Proceedings of IEEE International Symposium on Network Coding, 2013, pp. 1–5.

[2] P. Delsarte, *On subfield subcodes of modified Reed-Solomon codes*, IEEE Transactions on Information Theory, **21** (1975), no. 5, 575–576.

[3] C. Ding, and J. Yang, *Hamming weights in irreducible cyclic codes*, Discrete Mathematics, **313** (4) (2013), 434–446.

[4] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, *On the locality of codeword symbols*, IEEE Trans. Inform. Theory **58** (2011), no. 11, 6925–6934.

[5] S. Goparaju and R. Calderbank, *Binary cyclic codes that are locally repairable,* Proc. 2014 IEEE Int. Sympos. Inform. Theory, Honolulu, HI, pp. 676–680.

[6] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, *Codes with local regeneration and erasure correction*, IEEE Transactions on Information Theory **60** (8) (2014), pp. 4637–4660.

[7] R. J. McEliece, *Irreducible cyclic codes and Gauss sums*, Combinatorics, 1975, pp. 185–202.

[8] D.S. Papailiopoulos and A.G. Dimakis, *Locally repairable codes*, Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, 2012, pp. 2771–2775.

[9] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, *Optimal linear codes with a local-error-correction property*, Proc. 2012 IEEE Internat. Sympos. Inform. Theory, IEEE, 2012, pp. 2776–2780.

[10] N. Silberstein, A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath, *Optimal locally repairable codes via rank-metric codes*, arXiv:1301.6331.

[11] I. Tamo and A. Barg, *A family of optimal locally recoverable codes*, IEEE Transactions on Information Theory **60** (2014), no. 8, 4661–4676.

[12] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, *Optimal locally repairable codes and connections to matroid theory*, Proc. 2013 IEEE Internat. Sympos. Inform. Theory, 2013, pp. 1814–1818.

[13] J. H. van Lint, *Introduction to coding theory*, Vol. 86, Springer-Verlag, Berlin, 1992.

[14] A. Wang and Z. Zhang, "An integer programming based bound for locally repairable codes," arXiv:1409.0952.

APPENDIX A
BOUNDS ON THE DISTANCE OF LRC CODES

In the examples in Section IV we construct a number of examples of LRC codes over small alphabets (binary, and in one example, ternary). To assess how far the constructions are from being distance-optimal, we use upper bounds as a proxy for optimality. In this section we collect some of the upper bounds on the distance of codes with locality

Apart from the *Singleton-like bound* mentioned above and its refinements (e.g., [14]), the following two upper bounds on the cardinality of a q-ary (n,k,r) LRC code are known. A *shortening* bound was proved in [1]. We formulate it for the case of linear codes. Let $k_q(n,d)$ be the largest possible dimension of a linear q-ary code of length n and distance d. The maximum dimension $\mathcal{K}(n,r,d)$ of a q-ary linear LRC code of length n, distance d, and locality r satisfies the following inequality:

$$\mathcal{K}(n,r,d) \le \min_{1\le t\le \nu}(tr + k_q(n-t(r+1),d)). \quad (11)$$

If the code $\mathcal{C}$ is cyclic, then obviously the condition that the locality is r is equivalent to the condition that the dual distance $d^{\perp} := d(\mathcal{C}^{\perp}) = r+1$. Denote by $M_q^{(c)}(n,r,d)$ the maximum cardinality of a cyclic q-ary code of length n, locality r, and distance d. We can use the following form of the Delsarte *linear programming bound* [13] on the largest possible size of a q-ary cyclic LRC code of length n and locality r: $\mathcal{C}$ with distance d :

$$\begin{aligned} M_q^{(c)}(n,r,d) \le 1 + \max\Big\{ &\sum_{i=d}^{n} a_i : a_i \ge 0, i = d,\dots,n,\\ &\sum_{i=d}^{n} a_i K_k(i) = -\binom{n}{k}(q-1)^k, k = 1,\dots,r+1,\\ &\sum_{i=d}^{n} a_i K_k(i) \ge -\binom{n}{k}(q-1)^k, k = r+2,\dots,n\Big\}, \quad (12)\end{aligned}$$

where $K_k(i)$ is the value of the Krawtchouk polynomial of degree k. The question of the goodness of the bounds (11), (12) is currently very much open, and there is a gap between them and the parameters of many codes in examples.

APPENDIX B
PROOF OF PROPOSITION 4.4

Let $\mathbb{F}_{q^z}$ be a subfield of $\mathbb{F}_{q^m}$ and let $T_{m/z} := T_{\mathbb{F}_{q^m}/\mathbb{F}_{q^z}}$ be the trace mapping from $\mathbb{F}_{q^m}$ to $\mathbb{F}_{q^z}$. We abbreviate $T_{\mathbb{F}_{q^m}/\mathbb{F}_q}$ as T_m.

Define the subspace

$$V_z = \{(T_z(\gamma), ..., T_z(\gamma\beta^{s-1})), \gamma \in \mathbb{F}_{q^z}\},$$

where $z = \mathrm{ord}_s(q)$, and β is an s-th primitive root of unity.

Similarly define

$$V_m = \{(T_m(\gamma), ..., T_m(\gamma\beta^{s-1})) : \gamma \in \mathbb{F}_{q^m}\},$$

We will prove that $V_m = V_z$.

Proof that $V_m \subseteq V_z$. Let $(T_m(\gamma), ..., T_m(\gamma\beta^{s-1})) \in V_m$ for $\gamma \in \mathbb{F}_{q^m}$. Recall that $T_m = T_z \circ T_{m/z}$. We have

$$\begin{aligned} &(T_m(\gamma),\dots,T_m(\gamma\beta^{s-1}))\\ &= (T_z(T_{m/z}(\gamma)),\dots,T_z(T_{m/z}(\gamma\beta^{s-1})))\\ &= (T_z(T_{m/z}(\gamma)),\dots,T_z(T_{m/z}(\gamma)\beta^{s-1})) \in V_z. \end{aligned}$$

Proof that $V_m \supseteq V_z$. Since $T_{m/z}$ is surjective, there exists $\gamma' \in \mathbb{F}_{q^m}$ such that $T_{m/z}(\gamma') = \alpha \in \mathbb{F}_{q^z}\backslash\{0\}$. Let $(T_z(\delta), ..., T_z(\delta\beta^{s-1})) \in V_z$ for $\delta \in \mathbb{F}_{q^z}$. We show that this vector belongs also to V_m. Consider the following vector in V_m :

$$\Big(T_m\Big(\frac{\gamma'\delta}{\alpha}\Big), ..., T_m\Big(\frac{\gamma'\delta}{\alpha}\beta^{s-1}\Big)\Big), \text{ for } \frac{\gamma'\delta}{\alpha} \in \mathbb{F}_{q^m}.$$

Then

$$\begin{aligned} &\Big(T_m\Big(\frac{\gamma'\delta}{\alpha}\Big),\dots,T_m\Big(\frac{\gamma'\delta}{\alpha}\beta^{s-1}\Big)\Big)\\ &= \Big(T_z\Big(T_{m/z}\Big(\frac{\gamma'\delta}{\alpha}\Big)\Big),\dots,T_z\Big(T_{m/z}\Big(\frac{\gamma'\delta}{\alpha}\beta^{s-1}\Big)\Big)\Big)\\ &= \Big(T_z(\frac{\delta}{\alpha}T_{m/z}(\gamma')),\dots,T_z(\frac{\delta\beta^{s-1}}{\alpha}T_{m/z}(\gamma')))\\ &= \Big(T_z\Big(\frac{\delta}{\alpha}\alpha\Big),\dots,T_z\Big(\frac{\delta\beta^{s-1}}{\alpha}\alpha\Big)\Big)\\ &= (T_z(\delta),\dots,T_z(\delta\beta^{s-1})), \end{aligned}$$

and the result follows. The rest of the proof follows from Theorem 4.3.